DOCUMENT RESUME

ED 231 844 TM 830 312

AUTHOR Dole, Arthur A.; And Others
TITLE Meta-Analysis of Outcome Research in Reducing Test Anxiety: Interventions, Rigor, and Inertia.
PUB DATE Apr 83
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Counseling; Higher Education; *Intervention; Rating Scales; *Research Design; *Self Evaluation (Individuals); *Test Anxiety; Test Construction; *Undergraduate Students
IDENTIFIERS Effect Size; *Inertia Paradigm; *Meta Analysis; Rigor (Evaluation)

ABSTRACT

A meta-analysis of 78 published studies, from 1966 to 1979, on test anxiety reduction in college students was concerned with: (1) to what extent did treatment methods differ in efficacy; (2) to what extent was rigor in experimentation associated with outcome; and (3) to what extent was inertia, as a characteristic of research, associated with outcome? Each study was rated; the values on the measures of Rigor and Inertia were audited and discrepancies adjusted. Effect Size was computed, interventions identified, and the reliability of these procedures was established. Preliminary analyses of both the DiTomasso Methodology Inventory (DMI) and the Checklist for Ert vs. Inert Qualities (CEIQ) included item factor analyses, Cohen's Kappa of individual items, item means, and standard deviations. The average Effect Size--for each study for which information was available--and Rated Importance were correlated with Rigor as measured with the DMI, and Inertia, as measured by the CEIQ. In general, the findings were consistent with those of Glass and associates in confirming the value of contemporary counseling and psychotherapy as an intervention for college students with self-reported test anxiety. It appears that simple relaxation is as helpful as more complex and time consuming treatments. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED231844

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Meta-analysis of Outcome Research in Reducing Test Anxiety: Interventions, Rigor, and Inertia

Arthur A. Dole
University of Pennsylvania

Patience B. Rockey
Beaver College

Robert DiTomasso
West Jersey Hospital

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. A. Dole

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper prepared for presentation at American Educational Research Association, April 11-15 1983. Montreal, Canada

Running Head: Reducing Test Anxiety

TM 830 312

Printed in U.S.A.

Meta-analysis of Outcome Research in Reducing Test Anxiety: Interventions, Rigor, and Inertia

Arthur A. Dole
University of Pennsylvania

Patience B. Rockey
Beaver College

Robert DiTomasso
West Jersey Hospital

In outcome studies when college students are assisted in reducing their anxiety about classroom tests, does the type of intervention or the rigor of the experiment make a difference? If a group of published studies are rated for inertia, that is, for adherence to a physical s cience research model, is this characteristic related to outcome? In this meta-analysis we have three major objectives: to verify the work of Smith and Glass (1977) as applied specifically to five methods of helping undergraduates to control their fears of tests, to test the hypothesis that the quality of experimental design is related to effect size, and finally to explore an "ert" approach (Dole and Rockey, 1982) as applied to contemporary research on test anxiety.

After a meta-analysis of approximately 400 published, controlled studies in psychotherapy, Glass concluded that several treatment methods were roughtly comparable in effect. Later in their book about the same meta-analysis, Smith, Glass,and Miller (1980) substantiated that those who received treatment with a variety of interventions for a variety of dysfunctions tended to benefit more than controls on a variety of outcome measures. Landman and Dawes (1982) then confirmed their findings for a stringently selected portion of Smith and Glass' sample.

In the present research we have applied meta-analysis to a subset limited to published studies in which the treatment was applied to college students with test anxiety. We were curious as to whether or not Glass' overall conclusions would be sustained for test anxiety reduction studies.

Among others, Tryon (1980) has called attention to the issue of rigor in test anxiety research. We reasoned that if the experimental rigor of the studies was associated with outcome, such a finding would contribute to the design of subsequent studies. One of us (DiTomasso, 1981) developed a rating scale for experimental rigor, defined as the extent to which threats to validity were incorporated in a given research as published.

On the other hand, test anxiety may be construed as a personal, idiosyncratic internal process or a series of private events. Studies which recognized the "ert" (Dole & Rockey, 1982) aspects of the research may differ also in outcome as contrasted with more inert studies. Two of us(Dole & Rockey, 1982; Dole & Rockey, 1978, Dole, Note 1) have developed an ert appro ch to human behavior,and for psychotherapy research in particular we developed a rating scale for published articles, which measures the extent of inertia. Thus, we saw in meta-analysis a possible method to explore some of our conceptions about inertia as a paradigm in psychological research.

Research Questions

Our research questions, then, concerned a group of published studies on test anxiety reduction in college students: (a) To what

extent did treatment methods differ in efficacy; (b) to what extent was rigor in experimentation associated with outcome, and (c) to what extent was inertia, as a characteristic of research, associated with outcome?

Method

Studies

Seventy-eight studies were identified and xeroxed for the meta-analyses. All were published from 1966 to 1979 in a refereed journal; the subjects were students from a wide variety of institutions; one or more interventions were used to reduce anxieties about classroom tests; control groups were sometimes used. Objective outcomes were apprasied by test anxiety inventories, grade point average, performance tests and so on. Single case studies were included. As cited comprehersively in DiTomasso (1981), the journals included the Journal of Abnormal Psychology, Behavior Therapy, Behavior Research Therapy, The Journal of Counseling Psychology, the Journal of Behavior Therapy and Experimental Psychiatry, the Journal of Consulting and Clinical Psychology and others. Although there was considerable overlap with the Smith, et al. (1980) meta-analyses, ours differed in that they were restricted to studies about reduction of test anxiety, they did not include dissertations, and they extended the time frame to 1979. The number of studies available for each analysis ranged from 46 to 78.

Outcome Measures

Two outcome measures were used, Effect Size and Rated Importance. For 46 studies with control groups it was possible to average one or more measures of Effect Size according to Smith and Glass' formula as adapted

by DiTomasso (1981). Thus an effect size was defined by taking the difference between the mean of the treatment group and the control group and dividing it by the standard deviation of the control group ($X_t - \bar{X}_c / S_c$, Smith & Glass, 1977). When information was incomplete (no standard deviations, no control group, etc.) effect sizes were calculated according to suggestions by Glass (1978). See DiTomasso (1981) for explanations in detail about how the formula was modified.

The underlying logic of effect size is based upon its characteristic as a standard score. As a form of a normal unit deviate, effect sizes have a mean of 0 and a standard deviation of 1; thus the proportion of clients in the control group who are worse off than the average client receiving therapy is immediately evident.

Since more than one effect size was available in each study in most instances, two strategies were used. First we calculated an <u>average</u> effect size for each investigation. Second, to analyze the effect of the various treatment we calculated effect size by dependent measures classified as test anxiety, trait anxiety, grade point average, and performance. By allowing more than one effect size from each study to enter separately into the data pool, undoubtedly we altered complicated patterns of unwanted dependency and errors and violated some assumptions of differential statistics. See DiTomasso (1981) for a fuller discussion of the advantages and disadvantages of these procedures.

Note that for 32 studies, including eight small case studies, it was not possible to apply a quantitative index for effect size. Thus, as a second measure we developed rated importance. For all studies two or more independent raters applied a

Likert 5-point scale of "clear practical or theoretical importance," ranging from 1, "important differences for no hypotheses or questions," to 5,"important differences on all hypotheses or questions." The trained raters (not the investigators) were asked to assume that "an important difference was one which has clear, practical or theoretical consequences." Since the raters failed to attain satisfactory congruence, we reviewed and resolved all differences between the ratings and considered this measure as suggestive rather than definitive. As indicated by an r of .11 with average Effect Size, the two outcome measures were independent.

Interventions

For 46 studies DiTomasso (1981) selected and classified five interventions for comparison: systematic desentization, variants of systematic desentization, relaxation therapy, study skills training, and other. On the basis of commonly accepted descriptions of the treatment packages, Wolpe's (1958, 1973, 1976) publications served as a resource for defining desentization, its variants, and relaxation training. Study skills interventions were defined simply as any treatments that focused mainly on remediating faulty study habits, skills, or attitudes of test anxious students. The remaining category (other) served as a general category for any additional treatment packages included within the study. The other category of therapies included cognitive therapy, covert positive reinforcement, psychodrama, implosion, hypnosis and multiple treatment. Independent raters showed satisfactory congruence in classifying the treatments (DiTomasso, 1981).

It is conceivable that the extent of therapist's experience, the duration of the treatment, the setting of the psychological services, level of academic standards in a specific school, the size of the school, and whether or not the school was private or public might have influenced the efficacy of treatment. However, DiTomasso (1981) found that none of these variables was significantly associated with average effect size and they are therefore not addressed in this report.

Rigor

The DiTomasso Methodology Inventory (DMI) is a 51 item measure of experimental quality. Typical items are, "Population from which the sample is selected is clearly specified," and, "Deliveries of all treatments are adequately checked to insure standardized presentation." Each item is scored pass or fail on the basis of the manual. The range in possible scores is from 0 to 51. After training, two independent judges rated each article; disagreements were resolved by one of the investigators. According to DiTomasso (1981), the instrument possesses content validity, construct validity and acceptable levels of item interjudge reliability, and alpha reliability.

Inertia

The Checklist for Ert vs. Inert Qualities (CEIQ) comprises 32 Likert 5 point scales. CEIQ is intended to measure the extent to which a research study conforms to a physical science model. Typical items include, "What relation does the writer assume between overt and covert behavior of subjects," and "To what extent does the design provide for a possible change in the subjects." Since by inspection only two

items appeared to resemble DMI items, the CEIQ was retained in its entirety. Two independent judges (trained by the authors and not those who rated the DMI) rated each article and then any differences in their ratings were resolved. The range of possible scores was from 32 to 160. To establish congruence the raters were given a set of 21 articles from recent issues of the Journal of Counseling Psychology, which reported outcome studies other than reducing test anxiety. Scores assigned by Coder 1 correlated .79 with coder 2. Means did not differ significantly. On the basis of item analyses, factor analyses, and congruence analysis by item (Cohen's kappa), we concluded that it was appropriate to use the total score and selected factor scores but not single items.

Procedure

After the articles were xeroxed and the judges had been trained on practice articles to follow scoring instructions, each published study was read carefully and was rated; the values on the measures of Rigor and Inertia were also audited and discrepancies adjusted. In addition one of us, DiTomasso (1981), computed Effect Size, identified interventions, and established the reliability of these procedures. Preliminary analyses of both the DMI and the CEIQ included item factor analyses, Cohen's kappa of individual items, item means and standard deviations. For all instances in which data were available for the five interventions, Effect Size means on test anxiety, trait anxiety, grade point, and performance were calculated. That is, the standardized Effect Size measures were treated separately according to the classification of the outcome measure. Then average Effect Size for each study for which

information was available and Rated Importance were correlated with Rigor as measured with the DMI and Inertia as measured by the CEIQ. In addition, on an exploratory basis, nine factor scores for Inertia were correlated with the other variables.

Results

Interventions

In general, counseling and psychotherapy appeared to produce moderately positive improvement in self-reported test anxiety among college students. Other methods, systematic desentization, variants of systematic desentization, and relaxation therapy appeared to be more effective than study skills.

According to Table 1, for all therapy classifications except study skills, the average Effect Size somewhat exceeded 1.00 on test anxiety and was substantially lower on trait anxiety. On grade point average and performance effect sizes ranged from slightly positive to moderately negative.

Insert Table 1 about here

Rigor

Measured Rigor of the experiment (DMI) correlated moderately and negatively ($r = -.38$, $p < .01$) with rated Importance of the findings and slightly and negatively with Effect Size ($r = -.18$, $p < .05$) (Table 2). Rigor also correlated positively with inertia (CEIQ) ($r = .45$, $p < .01$).

Insert Tables 2 and 3 about here

Inertia

The measure of Inertia (CEIQ) was not significantly associated with either outcome measure (Table 2). This finding was confirmed with a subset of 21 nontest anxiety articles from the Journal of Counseling Psychology ($r = -16$).

Nine inert factor scores were correlated with Effect Size, rated Importance, and Rigor for the test anxiety studies (Table 3). None was related to average Effect Size. Factor 1, scientism, was slightly negatively related to rated Importance of the study. Factor 1, Scientism ($r = .46$) and Factor 2, Expertise ($r = .40$) were moderately associated with the DMI. Table 4 reports the means, and standard deviations of all measures and Table 5, the factor analysis of CEIQ, with highest items loadings.

Insert Tables 4 and 5 about here.

Discussion

In general, our findings were consistent with those of Glass and associates in confirming the value of contemporary counseling and psychotherapy as an intervention for college students with self-reported test anxiety. Among a variety of practical and theoretical implications it appears that simple relaxation is as helpful as more complex and time-consuming treatments. Consistent with literature reviews (Tryon, 1980; Emmelkamp, 1982), systematic desentization and its variants, relaxation, plus a large variety of other treatments, have been found to reduce test anxiety much more than untreated controls and in pretest-postest designs reported trait anxiety has also diminished substantially. On the other hand, in most studies academic performance, as measured by

grade point average or standardized achievement tests, either did not change or rose only moderately after treatment.

The quality of methodology as rated in this meta-analysis tended to be moderately associated with outcomes which were judged less important. Among possible explanations: Relatively poor studies may have inflated outcomes, tight methodologies may have "lost" important effects, and more rigorous investigators may have overlooked or excluded certain possible consequences of treatment. Since average Effect Size was independent of rated methodological Rigor and rated Importance, we speculate that in this group of research studies on reducing test anxiety the pattern of methodological flaws was such that there was an equal likelihood that between group differences were spuriously inflated or restricted. These findings would seem to strengthen the case for meta-analyses on the basis that false effects will tend to randomize.

Our expectation that, when inert paradigms are applied to studies of anxiety, outcomes would be reduced in size or importance, was not confirmed. However, this finding may be in part a function of constriction in range. That is, those studies which were published in refereed journals for the period under investigation were predominantly inert. In our exploratory factorial study of the Ert vs. Inert Checklist we found a rather complex factor structure. However, Scientism was slightly but significantly negatively related to rated Importance.

We have maintained that to a considerable extent the experience of test anxiety includes characteristics that are more ert than inert. For a particular college student a series of cognitions and affects, which are fluid, private and internal, accompany what happens before,

during and after an examination. The personal history of the student, the peculiarities of the instructor, and his or her pedagogical methods, influence test anxiety.

As possible ert approaches to subsequent research we suggest single case designs, custom-tailored interventions and projective or non-obtrusive outcome measures as supplements to conventional designs. There is certainly nothing in this research which would support the notion that such ert approaches would increase or decrease the rated importance of the study or effect size on standard outcome measures.

Finally, we suggest an ert approach to professional practice in reducing test anxiety. We are thinking of self-regulation, personalized hierarchies in desentization, collaboratively constructed guided imagery, and other individualized treatment strategies. We also suggest that professional practitioners adopt a questioning approach (Dole and Rockey, 1982) in reducing test anxiety.

Reference Note

1. Dole, A. A. Change within higher education: Some ert conceptions. Paper presented at the meeting of the International Council of Psychologists, Southampton, England, July 1982.

References

Dole, A. A., & Rockey, P. B. Toward rigorous but ert models of man. Journal of Religion and Health, 1977, 16, 81-90.

Dole, A. A., & Rockey, P. B. Ert approaches to research design. Academic Psychology Bulletin, 1982, 4, 335-344.

DiTomasso, R. A. A meta-analysis of test anxious college students. Unpublished doctoral dissertation, University of Pennsylvania, 1981.

Emmelkamp, P. M. G. Anxiety and fear. In A. S. Bellak, M. Hersen, and A. E. Kazdin (Eds.) International Handbook of Behavior Modification and Therapy. New York: Plenum, 1982.

Landman, J. T., & Dawes, R. M. Psychotherapy outcome: Smith and Glass conclusions stand up under scrutiny. American Psychologist, 1982, 37, 504-516.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Smith, M. L., Glass, G. V., & Miller, T. T. The benefits of psychotherapy. Baltimore, Md.: Johns Hopkins University Press, 1980.

Tyron, G. S. The measurement and treatment of test anxiety. Review of Educational Research, 1980, 50, 343-372.

Wolpe, J. Psychotherapy by reciprocal inhibition. California: Stanford University Press, 1958.

Wolpe, J. The practice of behavior therapy. New York: Pergamon, 1973.

Wolpe, J. Theme and variations: A behavior therapy casebook. New York: Pergamon, 1976.

Footnotes

1. We acknowledge the assistance of Maria Showalter in analyzing the date for tnis meta-analysis. Helene Godoroff assisted in coding the CEIQ.

Table 1

Average Standard Effect Sizes After Various Types of Therapy by Measure

Therapy	Measure	No. Effect Sizes	No. Studies	Mean	Standard Deviation	Percentile Rank
Standard systematic	Test anxiety	3	1	-1.24	.34	11
Desensitization	Trait anxiety	0	0	----	---	--
	Grade point average	1	1	-.27	---	39
	Performance	3	1	-.01	.26	50
Desensitization	Test anxiety	71	27	-1.31	.80	10
Variants	Trait anxiety	27	10	-1.00	1.04	16
	Grade point average	15	8	+.50	.44	69
	Performance	55	15	+.23	.72	59
Relaxation therapy	Test anxiety	25	10	-1.18	.79	12
	Trait anxiety	14	6	-.65	1.07	26
	Grade point average	2	2	+.12	.48	55
	Performance	14	6	-.11	.87	55

Table 1 (Continued)

Average Standard Effect Sizes After Various Types of Therapy by Measure

Therapy	Measure	No. Effect Sizes	No. Studies	Mean	Standard Deviation	Percentile Rank
Study skills	Test anxiety	6	4	-.30	.43	38
Counseling	Trait anxiety	2	2	-.46	.13	32
	Grade point average	0	0	----	---	---
	Performance	7	2	+.29	1.43	61
Other	Test anxiety	34	15	-1.13	.47	13
	Trait anxiety	6	3	-1.45	1.33	7
	Grade point average	6	6	+1.03	.53	85
	Performance	26	9	+.54	1.26	71

Note: For each measure, percentile rank indicates the percentage of subjects in the control group in relation to the average treated person. Effect size is based on Smith and Glass (1977) as adapted by DiTomasso (1981).

Table 2

Correlations of Effect Size and Rated Importance
With Rigor and Inertia for Studies of Test Anxiety

	Rated Importance	Rigor	Inertia
Effect size[1]	.11	-.18	.11
	(46)	(46)	(46)
Rated importance[2]	---	-.38**	-.17
		(70)	(78)
Rigor		---	.45**
			(70)

*p < .05

**p < .01

1. Effect size based on Smith and Glass (1977) as adapted by DiTomasso (1981).
2. Importance was rated on a Likert scale from (1) "important differences for no hypotheses or research questions" to (5) "important differences on all hypotheses or questions."

Note: Number of studies is indicated in parentheses.

Table 3

Correlations of Inert Factors with Effect Size,

Rated Importance and Rigor for Test Anxiety Studies

		Effect Size[1] (N = 46)	Rated Importance[2] (N = 78)	Rigor (N = 70)
I	Scientism	-.14	-.22*	.46**
II	Expertise	-.11	.07	.40**
III	Objectivity	.08	-.07	.13
IV	Universality	-.09	.04	.00
V	Externality	-.01	.04	.14
VI	Generality	-.08	.14	-.21
VII	Multimethod	.08	.13	-.08
VIII	Independent Variables	.08	.18	-.16
IX	Deduction	.16	.01	.16

*$p < .05$

**$p < .01$

1. Effect size based on Smith and Glass (1977) as adapted by DiTomasso (1981).
2. Importance was rated on a Likert scale from (1) "important difference for no hypotheses or research questions" to (5) "important differences on all hypotheses or questions."

Table 4

Means and Standard Deviations on Dependent Measures

for Studies on Reducing Test Anxiety

Measure	N	Mean	S.D.
DMI	70	20.33	5.46
Inert Qualities	78	4.31	.38
Rated Importance	78	3.79	.99
Effect Size	46	1.14	.79
Factor I[1]	78	4.50	.65
Factor II	78	4.75	.51
Factor III	78	4.68	.41
Factor IV	78	4.16	.75
Factor V	78	2.39	.67
Factor VI	78	4.46	.43
Factor VII	78	3.44	.27
Factor VIII	78	4.82	.39
Factor IX	78	3.56	1.06

[1]Factor scores have been adjusted for unequal N's.

Table 5

Highest Factor Loading by Item:

Checklist for Inert vs. Ert Qualities

Factor I:	Scientism	Item	Loading
M-19	What characterizes the assumptions underlying statistics?		912
R-1	How are the differences recorded?		860
D-2	What relation does the writer assume between overt and covert behavior of subjects?		854
M-16	How are the data analyzed?		841
M-10	How are the psychometric properties of measures presented?		839
D-4	After intervention, if any, what do the findings about the participants imply?		791
R-3	How was the unaccounted variance reported?		790
R-4	What kind of implicit reasoning characterizes the report of results?		773
M-7	Which best describes the structure of data gathering?		728
M-2	How is a subject or participant described?		696
M-15	To what extent does the design provide for unique events?		673
S-3	The writer is concerned with the phenomenon at what level of abstraction (time and setting)?		660
D-3	To what extent does the writer assume the findings to be specific?		607
M-14	To what extent were the situations examined in the research controlled by the experimentor?		589
M-1	How many subjects or participants took part?		573
M-13	To what extent does the design provide for a possible change in subjects?		547

Table 5 (Continued)

Factor II:	Expertise	Item	Loading
M-8	Who conducts the data gathering?		873
D-1	Who interpreted the data?		837
Factor III:	Objectivity		
M-4	How is the experimentor described?		764
M-5	How was the participant involved in the study?		712
M-6	How are the circumstances described?		
Factor IV:	Universality		
S-2	To whom are the expressed purposes of the writer primarily applicable?		765
(M-13)	To what extent does the design provide for possible change in the subjects?		-454
Factor V:	Externality		
S-1	On which aspect of the topic does the writer focus?		802
M-11	So far as the independent variable(s) is concerned, on which does the writer tend to focus?		659
Factor VI:	Generality		
R-2	How are observations over time reported?		897
L-1	How is the review of the literature presented?		-412
Factor VII	Multimethod		
M-9	How many different techniques for data gathering were used?		762
M-12	So far as the dependent variable(s) is concerned, on which does the writer tend to focus?		-730